DOCUMENT RESUME

ED 250 394 TM 840 748

AUTHOR Arnaut, Lynn Y.; Bownas, David A.
TITLE The Effects of Rating Purpose and Anonymity on Self Ratings.
PUB DATE Mar 84
NOTE 10p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Adults; Evaluation Needs; *Job Performance; Personnel Evaluation; *Reliability; *Self Evaluation (Individuals)
IDENTIFIERS *Anonymity of Respondents; *Leniency Response Bias

ABSTRACT

The present study was designed to answer four questions: (1) do anonymity and purpose affect the leniency of self ratings in general, (2) do these factors interact, (3) do self ratings approximate supervisor ratings under certain conditions, and (4) is self-assessment a useful method of determining training needs? Self-ratings and supervisory ratings were made for 206 employees who were in one of three job categories in the United States Marshals Service. Subjects were told that ratings were for one of three purposes: training needs assessment, research, or personnel administration. In addition, they were instructed either to sign or not to sign their names. The results indicated that the stated purpose for self ratings had a significant effect on self rating leniency, and anonymity had no effect on self rating leniency. (BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED250394

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

The Effects of Rating Purpose and Anonymity on Self Ratings

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
L.Y. Arnaut
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Lynn Y. Arnaut

David A. Bownas

Virginia Polytechnic Institute
and State University
Blacksburg, Va.

TM 810 746

Employee self ratings are not widely used in training needs assessment, primarily because of the frequently cited finding of self rating leniency (Thornton,1980). However, self ratings have several advantages which make them useful for identifying training needs. First, they are economical. Second, individuals may have access to more information about their abilities and performance than do observers (Jones & Nisbett, 1971). Third, self evaluations typically show less halo than do supervisory ratings, which indicates that there is greater discrimination between rated factors for self ratings. If self rating leniency may be reduced, self ratings may be a useful method of identifying training needs.

The purpose for which the self ratings are made and the anonymity of the ratings are two factors which may influence self rating leniency (Landy & Farr, 1980; Mabe & West, 1982). With respect to anonymity, Mabe and West (1982) suggested that when self ratings are anonymous, the self-esteem of the rater would not be enhanced by high ratings. Thus, anonymous ratings should be less lenient than named ratings. Regarding purpose, if the purpose is such that an individual is likely to benefit from distortion, as in a selection or promotion procedure, then increased leniency is a potential problem.

In addition, anonymity and rating purpose may interact. When ratings are made for administrative reasons there is a high likelihood that spuriously high ratings may be viewed by an individual who can invalidate them. In this case, the individual may benefit more from accurate than from inflated ratings.

BEST COPY AVAILABLE

When the stated purpose is research rather than administration, the individual making a self-assessment is likely to derive little observable benefit from distorted ratings. However, ratings which can be identified may be inflated because heightened self-esteem may result for the rater. Even if the results can be checked for accuracy, which is an unlikely condition if the ratings are to be done for research purposes, the individual would not be penalized for inaccurate results.

Therefore, anonymous ratings made for research may show less leniency than named ratings, while confidential administrative ratings may exhibit more leniency than identifiable evaluations. Such effects may not typically appear since research ratings are usually anonymous or confidential while administrative ratings can usually be identified.

The present study was designed to answer four questions. First, do anonymity and purpose affect the leniency of self ratings in general? Second, do these factors interact as has been suggested above? Third, do self ratings approximate supervisor ratings under certain conditions? Fourth, is self-assessment a useful method of determining training needs?

Method

Subjects: Self-ratings were made by 206 employees who were in one of three job categories in the United States Marshals Service (USMS). Incumbents in these positions perform a variety of federal law enforcement tasks. In addition, each employee's immediate supervisor also rated the incumbent.

Measures: A rating booklet identifying 66 trainable

BEST COPY AVAILABLE

knowledges and skills which are important in the three jobs of interest was prepared. Incumbents' proficiency on each knowledge and skill was rated using a 9-point scale with anchors ranging from "inadequate" to "outstanding". In addition, overall performance was rated using the same scale.

Procedure: Incumbents were asked to complete one questionnaire indicating their proficiency on each of the 66 knowledges and skills and their overall performance level. Subjects were told that ratings were for one of three purposes: training needs assessment, research, or personnel administration. In addition, they were instructed either to sign or not to sign their names.

Incumbents work in one of 90 districts. Each district was randomly assigned to one of the six instruction conditions, and all employees in that district received the same set of instructions. A questionnaire was mailed to each incumbent with instructions from the Director of the USMS to return the booklets promptly. Supervisory questionnaires were mailed two weeks after the rating forms had been mailed to the incumbents. After both the self- and supervisory rating booklets had been returned and all analyses had been completed, a debriefing letter was sent to all employees and supervisors.

Results

Two dependent measures were used for all analyses except those for halo. The first measure was the overall performance rating, and the second was the mean knowledge and skill rating, averaged across the 66 items.

BEST COPY AVAILABLE

The first question of interest was whether or not anonymity and purpose affected self rating level. An analysis of variance indicated that there was a significant effect of rating purpose on leniency for both the performance and the average ability ratings ($F(2,169)=4.93$, $p=.008$; $F(2,188)=4.28$, $p=.015$, respectively). Anonymity did not have a significant effect on either the performance measure ($F(1,169)=1.89$, $p=.171$) or the average ability measure ($F(1,188)=1.04$, $p=.310$). There were no significant anonymity-by-purpose interactions for either the performance measure ($F(2,169)=1.22$, $p=.298$) or the average ability measure ($F(2,188)=0.97$, $p=.382$).

In order to determine which group differences were significant, Student Newman-Keuls paired comparison tests (Kirk, 1982) were performed for the purpose variable. For the performance rating, the administrative and research conditions did not differ significantly, while the mean of the performance ratings in the training group was significantly lower than both the administrative and research conditions ($p<.05$). No significant pairwise differences were obtained for the average ability rating.

Although the six-group anonymity-by-purpose interaction was not significant, the second question concerned a possible interaction between anonymity and purpose for the research and administrative groups only. The hypothesized four-group interaction was not obtained for the ability measure; thus, no comparison was performed for this measure. For the performance measure the four-group interaction was in the predicted

BEST COPY AVAILABLE

direction. A planned comparison indicated that the differences between the groups were not significant ($F(1,169)=2.81$, $p=.10$).

The third question of interest was whether self ratings can approximate supervisory ratings under certain conditions. Overall, the mean self rating for both the performance measure and the average ability rating were significantly higher than the mean supervisory ratings ($t'(156)=3.67$, $p=.0005$ for performance; $t'(169)=3.89$, $p=.0005$ for average ability). The overall correlation between the self and supervisor performance ratings was only .04 ($p=.60$), and the overall correlation between the average ability measures was .12 ($p=.12$). These correlations did not vary when looked at as a function of instructional condition.

The final question concerned the use of self ratings for training needs assessment. In addition to the leniency of the ratings, the amount of halo present is also relevant. For the halo analysis, only items which met two criteria were included. First, an item had to be equally important for all three jobs. Second, items selected for the halo analysis were chosen to represent conceptually different categories of knowledges and skills, thus representing separate dimensions of performance. One operational definition of halo requires an analysis of the item intercorrelations. The average z' coefficient for the supervisory ratings was 0.673, and the average z' coefficient for the self ratings was 0.365. This difference was significant, ($z=2.81$, $p=.003$), indicating lower halo for self than for supervisory ratings.

A second operational definition assumes that restricted

BEST COPY AVAILABLE

variability across dimensions is due to a halo effect. The supervisor ratings showed a significantly smaller average standard deviation across rated abilities than the self ratings ($t'(169)=2.17$, $p=.025$). Thus, using both definitions of halo, the self ratings showed a significantly smaller halo effect than the supervisor ratings.

Looking at self ratings alone, the correlational definition of halo did not indicate differences in the amount of halo exhibited across the six treatment groups. Using the standard deviation definition of halo, the purpose manipulation had a significant effect on self rating halo; a paired comparison indicated that the training group had significantly less halo than the other two groups ($p<.05$).

The variance of the supervisory ratings was significantly greater than the self rating variance for both dependent measures ($F\text{-max}(2,169)=3.99$, $p<.01$ for the performance measure; $F\text{-max}(2,169)=2.58$, $p<.01$ for the average ability rating).

Conclusions

With respect to purpose, the results indicated that the stated purpose for self ratings had a significant effect on self rating leniency for both dependent measures. When subjects were told that the purpose of the ratings was training needs assessment, both dependent measures were less inflated than when the stated purpose was research or administration, although this difference was only significant for the performance measure.

Anonymity had no effect on self rating leniency for either dependent measure. There are three possible reasons for this

BEST COPY AVAILABLE

lack of an effect. First, the subjects may not have believed that the questionnaires were in fact anonymous. A second possibility is that due to the nature of their jobs, the incumbents are simply more cautious than most populations in filling out questionnaires. A third possibility is that the performance and ability measures used here are less sensitive to anonymity manipulations than are the attitude measures for which anonymity effects have been reported (cf. Gordon & Petty, 1971). The failure to find the predicted four-group interaction indicates that the effect of rating purpose on leniency is not dependent upon the identifiability of the ratings.

The self ratings showed less halo than the supervisor ratings using both correlational and inter-item standard deviation operational definitions of halo. In addition, using the standard deviation operational definition, the training self ratings showed the least halo of the three purpose conditions.

Self-ratings exhibited less variance than supervisor ratings. In addition, for both dependent measures, self ratings were found to be more lenient than the corresponding supervisor ratings. However, given that the training self ratings were less lenient than the research and administrative ratings, and that self ratings in general (and training self ratings in particular) showed less halo than did the supervisor ratings, self ratings may indeed be a useful method of determining training needs.

BEST COPY AVAILABLE

References

Gordon, M.F. & Pet .y, M.M. A note on the effectiveness of research conditions in reducing the magnitude of dissimulation on a self-report criterion. Personnel Psychology, 1971, 24, 53-61.

Jones, E.F. & Nisbett, R.E. The actor and the observer: Divergent perceptions in the causes of behavior. New York: General Learning Press, 1971.

Kirk, R.E. Experimental Design: Procedures for the Behavioral Sciences. (2nd ed.). Belmont, Ca.: Brooks/Cole Publ. Co., 1982.

Landy, F.J. & Farr, J.L. Performance rating. Psychological Bulletin, 1980, 87, 72-107.

Mabe, P.A., III, & West, S.G. Validity of self evaluation of ability: a review and meta-analysis. Journal of Applied Psychology, 1982, 67, 280-296.

Thornton, G.C., III. Psychometric properties of self-appraisals of job performance. Personnel Psychology, 1980, 33, 263-271.

BEST COPY AVAILABLE